

Log Governance Dev Design

1. Config

kylin-server-log4j.xml

- Add a new file kylin.build.log to output some of the build-related logs, which all build logs will be added to this file later
- Add a new file kylin.metadata.log to output metadata and transaction related logs
- Adjust the level of Spark logs related to query logs
 - org.apache.spark.scheduler.TaskSetManager: Info -> WARN
 - org.apache.spark.scheduler.DAGScheduler: Info -> WARN
 - org.apache.spark.scheduler.YarnScheduler: Info -> WARN
- Remove duplicate log settings at the org.apache.kylin.query level

2. Query

1. QueryCacheManager

- Change log level of SQL with cache key from Info to Debug
- Change log level of the clearProjectCache from Debug to Info
- Update class QueryCacheManager log level to Info from Debug
- Log file is still kylin.query.log

2. org.apache.kylin.rest.service.QueryService

- Add a log marking the start of the query in kylin.query.log .

e.g.

```
1 Start query in project: xxx
2 ...
```

- Add a log marking the start of the async query in `kylin.query.log`.

e.g.

```
1 This query is an async query in project: xxx
2 ...
```

- Update of the timeline of the context metrics to align the query time and the sum of the individual sub-step times

```
1 QueryContext.current().record("update_metrics_time");
```

- Log file is still `kylin.query.log`

3. Build Log

- Build job schedule stage: `kylin.schedule.log`
- Execute stage: in dir `{hdfs_working_dir}/{project}/job_tmp`

e.g.: `{package_name}/job_tmp/{jobID}/01/execute_output.json.{timestamp}.log`

- Spring service log for build job: `kylin.build.log`

1. `org.apache.kylin.common.CustomUtils`

- New tool class, to output log in `kylin.build.log`

2. `org.apache.kylin.engine.spark.job.RDSegmentBuildExec`

- Update Detect source size duration calculation logic

3. Add log in stage to mark the beginning and end

```
1 Start sub stage xxx
2 ...
3 End sub stage xxx
```

- org.apache.kylin.engine.spark.job.stage.tablesampling.AnalyzerTable
- org.apache.kylin.engine.spark.job.stage.build.BuildDict
- org.apache.kylin.engine.spark.job.stage.build.BuildLayer
- org.apache.kylin.engine.spark.job.stage.build.FlatTableAndDictBase
- org.apache.kylin.engine.spark.job.stage.build.GatherFlatTableStats
- org.apache.kylin.engine.spark.job.stage.build.GenerateFlatTable
- org.apache.kylin.engine.spark.job.stage.build.MaterializedFactTableView
- org.apache.kylin.engine.spark.job.stage.merge.MergeColumnBytes
- org.apache.kylin.engine.spark.job.stage.merge.MergeFlatTable
- org.apache.kylin.engine.spark.job.stage.merge.MergeIndices
- org.apache.kylin.engine.spark.job.stage.merge.MergeStageTest.MergeStageMock
- org.apache.kylin.engine.spark.job.stage.build.partition.PartitionBuildDict
- org.apache.kylin.engine.spark.job.stage.build.partition.PartitionBuildLayer
- org.apache.kylin.engine.spark.job.stage.build.partition.PartitionGatherFlatTableStats
- org.apache.kylin.engine.spark.job.stage.build.partition.PartitionGenerateFlatTable
- org.apache.kylin.engine.spark.job.stage.build.partition.PartitionMaterializedFactTableView
- org.apache.kylin.engine.spark.job.stage.merge.partition.PartitionMergeColumnBytes
- org.apache.kylin.engine.spark.job.stage.merge.partition.PartitionMergeFlatTable
- org.apache.kylin.engine.spark.job.stage.merge.partition.PartitionMergeIndices
- org.apache.kylin.engine.spark.job.stage.build.partition.PartitionRefreshColumnBytes
- org.apache.kylin.engine.spark.job.stage.build.RefreshColumnBytes
- org.apache.kylin.engine.spark.job.stage.build.RefreshSnapshots
- org.apache.kylin.engine.spark.job.stage.snapshots.SnapshotsBuild
- org.apache.kylin.engine.spark.job.stage.WaitForResource

4. org.apache.kylin.engine.spark.builder.SegmentFlatTable

- Update output some log in kylin.build.log from kylin.log
 - Wrap ALIAS \${originDS.schema.treeString} TO \${newDS.schema.treeString}

- Lookup table schema \${lookupDataset.schema.treeString}

4. Metadata

1. KylinLogTool

- Add new file kylin.metadata.log in kylinLogPrefix

2. org.apache.kylin.rest.service.AuditLogService

- Output log "Start/End to catchup manually" to kylin.metadata.log from kylin.log

3. kylin.log -> kylin.metadata.log

- Output log to kylin.metadata.log from kylin.log in this class:
- org.apache.kylin.common.persistence.transaction.UnitOfWork
- io.kyligence.kap.metadata.epoch.EpochManager
- org.apache.kylin.common.persistence.transaction.AbstractAuditLogReplayWorker
- org.apache.kylin.common.persistence.transaction.AuditLogReplayWorker

5. Other

1. org.apache.kylin.rest.aspect.SchedulerEnhancer

- Update log "schedule at job leader" level is debug
- Update class SchedulerEnhancer log level is Info

2. org.apache.kylin.rest.service.task.QueryHistoryTaskScheduler

- Update construction method and method to thread scheduling to kylin.schedule.log from kylin.log

3. org.apache.kylin.metadata.cube.storage.TotalStorageCollector

- Update log "Reuse workingDirCapacity by project {}, storageSize: {}" level is debug
- Update class TotalStorageCollector log level is info

4. org.apache.kylin.common.metrics.MetricsInfluxdbReporter

- Update log level is debug:

"Start to aggregate daily metrics ..."

"Aggregate daily metrics success ..."

- Update class MetricsInfluxdbReporter log level is info

5. org.apache.kylin.engine.spark.source.NSparkMetadataExplorer

- Update permission check exception log level is error

6. org.apache.kylin.rest.broadcaster.Broadcaster

- Update level debug log and method consumeEvent 's log to kylin.schedule.log from kylin.log

7. io.kyligence.kap.metadata.recommendation.ref.OptRecV2

- Update log "initialize recommendation" level is debug
- Update class OptRecV2 log level is info

8. org.apache.kylin.rest.security.LdapAuthenticationProvider

- Update class LdapAuthenticationProvider log level is info

9. org.apache.kylin.rest.security.KylinAclPermissionEvaluator

- Update class KylinAclPermissionEvaluator log level is info

10. `org.apache.kylin.common.scheduler.EventBusFactory`

- Update log to `kylin.schedule.log` from `kylin.log`